Appl. Statist. (2000)
49, *Part* 4, *pp.* 441–451

Simple principal components

S. K. Vines

The Open University, Milton Keynes, UK

[Received April 1999. Revised February 2000]

Summary. We introduce an algorithm for producing simple approximate principal components directly from a variance–covariance matrix. At the heart of the algorithm is a series of 'simplicity preserving' linear transformations. Each transformation seeks a direction within a two-dimensional subspace that has maximum variance. However, the choice of directions is limited so that the direction can be represented by a vector of integers whenever the subspace can also be represented by vectors of integers. The resulting approximate components can therefore always be represented by integers. Furthermore the elements of these integer vectors are often small, particularly for the first few components. We demonstrate the performance of this algorithm on two data sets and show that good approximations to the principal components that are also clearly simple and interpretable can result.

Keywords: Interpretation; Pairwise linear transformation; Principal components analysis; Simplification

1. Introduction

Eigenvectors crop up in many multivariate techniques (see, for example, Krzanowski (1971)). In particular it is well known that the eigenvectors of a variance–covariance matrix correspond to the principal components for those data. In other words the eigenvector with the kth-highest eigenvalue defines the direction in which those data have the greatest variance, orthogonal to the directions given by eigenvectors with higher eigenvalues.

The widespread use of software that incorporates algorithms to evaluate eigenvectors means that principal components are easily obtained for data summarized by variance–covariance matrices. Unfortunately the precision of these algorithms hampers the interpretation of the directions once obtained. For example, the four principal components of the variance–covariance matrix for some data representing four separate measurements of the resistance index (RI) (a measure of resistance to flow in blood-vessels) in 444 pregnant women are

component 1,	0.423,	0.425,	0.548,	0.583,
component 2,	0.319,	−0.296,	0.649,	−0.625,
component 3,	0.579,	0.551,	−0.433,	−0.415,
component 4,	0.619,	−0.654,	−0.302,	0.311,

explaining 58%, 26%, 10% and 6% of the variance respectively.

Rules of thumb have developed for approximating the directions by more interpretable directions. For example, drastic rounding may be employed (Jackson, 1991) and/or elements of a principal component less than 70% in magnitude of the highest element may be set to 0

Address for correspondence: S. K. Vines, Department of Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK.
E-mail: s.k.vines@open.ac.uk

 0035–9254/00/49441

(Jeffers, 1967). However, such rules rarely result in orthogonality of the simplified components in addition to needing considerable skill to be applied well (see, for example, Cadima and Jolliffe (1995)). Alternatively the principal components may be rotated to try to produce a simpler and more interpretable set of components. However, rotations based on criteria such as the varimax criterion (Kaiser, 1958) tend to share out variance equally among components, losing the ordering of the components based on the variance explained.

Given the difficulty of simplifying principal components once obtained, we describe a different approach to obtaining simplified components in the next section — the generation of simple components directly from the variance–covariance matrix. We do this by considering a series of linear transformations to a set of simple orthogonal axes. In each transformation, the variance of the data with respect to one of the transformed axes is maximized. However, the transformation is restricted so that transformed axes can still be represented by simple directions, a simple direction being defined as a direction that can be represented by a vector proportional to an integer vector. Furthermore the transformed axis for which the data have the greater variance will tend to be proportional to a vector of small integers and hence be particularly simple. When only trivial transformations can be performed the final transformed axes are, by construction, simple directions which have properties close to the principal components. In other words the kth direction (the kth simple component) is a direction in which the data have a large variance orthogonal to the simple components with higher variances. Furthermore the correlation between two simple components will tend to be small.

For example we shall show that one set of four simple components obtained for the RI data is

component 1,	1,	1,	1,	1,
component 2,	1,	−1,	1,	−1,
component 3,	1,	1,	−1,	−1,
component 4,	1,	−1,	−1,	1,

explaining 57%, 24%, 11% and 9% of the variance respectively. Furthermore the correlation between the first two simple components, which together explain 81% of the total variance, is only 0.014.

We shall compare simple components generated by using our algorithm with the corresponding principal components for another data set in Section 3. Finally in Section 4 we shall draw some conclusions.

2. Generating simple components

2.1. A simplicity preserving transformation

Initially we consider a single linear transformation affecting just a couple of orthogonal directions $\mathbf{d}_1$ and $\mathbf{d}_2$ in p-dimensional space. This transformation is constructed so that orthogonality of the directions will be preserved although not necessarily their size. Hence it will be equivalent to orthogonally rotating and rescaling $\mathbf{d}_1$ and $\mathbf{d}_2$. In general such transformations can be written as $(\mathbf{f}_1, \mathbf{f}_2) = (\mathbf{d}_1, \mathbf{d}_2)P$ where

$$P = \begin{pmatrix} 1 & l_2^2\beta \\ \beta & -l_1^2 \end{pmatrix},$$

$l_1^2 = \mathbf{d}_1^{\mathrm{T}}\mathbf{d}_1$ and $l_2^2 = \mathbf{d}_2^{\mathrm{T}}\mathbf{d}_2$. Furthermore the variance–covariance matrix V^* of any data with respect to the new axes $\mathbf{f}_1$ and $\mathbf{f}_2$ is simply $P^{\mathrm{T}}VP$ where V is the variance–covariance matrix of

the data with respect to the old axes $\mathbf{d}_1$ and $\mathbf{d}_2$. In particular the variance v in the normalized direction $\mathbf{f}_1/l_{\mathbf{f}_1}$ can be written as

$$v = \frac{l_1^2 v_{11} + 2\beta l_1 l_2 v_{12} + \beta^2 l_2^2 v_{22}}{l_1^2 + \beta^2 l_2^2}$$

where

$$V = \begin{pmatrix} l_1^2 v_{11} & l_1 l_2 v_{12} \\ l_1 l_2 v_{12} & l_2^2 v_{22} \end{pmatrix}$$

and $l_{\mathbf{f}_1}^2 = \mathbf{f}_1^{\mathrm{T}} \mathbf{f}_1$. It can be shown that setting $\beta = \beta^*$ where

$$\beta^* = \begin{cases} \dfrac{-l_1 l_2 (v_{11} - v_{22}) + \sqrt{\{l_1^2 l_2^2 (v_{11} - v_{22})^2 + 4 l_1^2 l_2^2 v_{12}^2\}}}{2 l_2^2 v_{12}} & v_{12} \neq 0, \\ 0 & v_{12} = 0,\ v_{11} \geqslant v_{22}, \\ \infty & v_{12} = 0,\ v_{11} < v_{22} \end{cases}$$

maximizes v. Furthermore with this choice of β^* the data become uncorrelated with respect to the directions $\mathbf{f}_1$ and $\mathbf{f}_2$.

Unfortunately rewriting $\mathbf{f}_1$ as $\mathbf{f}_1 = \mathbf{d}_1 + \beta^* \mathbf{d}_2$ it is clear that, in general, $\mathbf{f}_1$ will not be a vector of integers even if $\mathbf{d}_1$ and $\mathbf{d}_2$ are. So, if we define a simple direction as a direction that can be written as (proportional to) a vector of integers, the transformation with $\beta = \beta^*$ does not preserve simplicity of directions.

We overcome this problem by restricting the values that β can take so that $\beta = i/2^k$ or $\beta = 2^k/i$, $i = -2^k, -2^k + 1, \ldots, 2^k$. For example when $k = 1$ this corresponds to the eight possible directions $\mathbf{d}_1 - 2\mathbf{d}_2$, $\mathbf{d}_1 - \mathbf{d}_2$, $\mathbf{d}_1 - \mathbf{d}_2/2$, $\mathbf{d}_1$, $\mathbf{d}_1 + \mathbf{d}_2/2$, $\mathbf{d}_1 + \mathbf{d}_2$, $\mathbf{d}_1 + 2\mathbf{d}_2$ and $\mathbf{d}_2$. The value $b^* = \beta$ which maximizes v is then sought only within this restricted set.

Setting

$$\left.\begin{aligned} \mathbf{d}_1^* &= \mathbf{f}_1 = 2^k \mathbf{d}_1 + 2^k b^* \mathbf{d}_2 \\ \mathbf{d}_2^* &= \mathbf{f}_2 = 2^k b^* l_2^2 \mathbf{d}_1 - 2^k l_1^2 \mathbf{d}_2 \end{aligned}\right\} \qquad |b^*| \leqslant 1,$$

$$\left.\begin{aligned} \mathbf{d}_1^* &= \mathbf{f}_1 = 2^k \mathbf{d}_1 / b^* + 2^k \mathbf{d}_2 \\ \mathbf{d}_2^* &= \mathbf{f}_2 = 2^k l_2^2 \mathbf{d}_1 - 2^k l_1^2 \mathbf{d}_2 / b^* \end{aligned}\right\} \qquad |b^*| > 1$$

ensures that $\mathbf{d}_1^*$ and $\mathbf{d}_2^*$ will be integer vectors whenever $\mathbf{d}_1$ and $\mathbf{d}_2$ are integer vectors. Hence simplicity of $\mathbf{d}_1$ and $\mathbf{d}_2$ under the transformation is preserved, albeit with a submaximal variance in the direction of $\mathbf{d}_1^*/l_1^*$ and some correlation of the data with respect to $\mathbf{d}_1^*$ and $\mathbf{d}_2^*$. Furthermore $\mathbf{d}_1^*$ will tend to be a simpler vector than $\mathbf{d}_2^*$ as its elements will, in general, be of lesser magnitude than the elements of $\mathbf{d}_2^*$.

2.2. Combining simplicity preserving transformations

We now assume that our p-dimensional data are described with respect to p orthogonal simple directions $\mathbf{d}_0(1), \ldots, \mathbf{d}_0(p)$. We also assume that the (r, s)th element $V_0(r, s)$, $r = 1, \ldots, p$, $s = 1, \ldots, p$, of the corresponding variance–covariance matrix V_0 for these data can be written such that

$$V_0(r, s) = l_{0_r} l_{0_s} v_0(r, s)$$

where l_{0_r} and l_{0_s} are the lengths of the rth and sth directions $\mathbf{d}_0(r)$ and $\mathbf{d}_0(s)$ respectively and $v_0(r, s)$ is the covariance of the data with respect to the normalized directions $\mathbf{d}_0(r)/l_{0_r}$ and $\mathbf{d}_0(s)/l_{0_s}$.

Consider a sequence of simplicity preserving transformations P_1, P_2,. . . as described in Section 2.1 applied to these data. It is easily shown that the variance–covariance matrix of the data after the ith transformation $V_i = P_i^{\mathrm{T}} V_{i-1} P_i$, $i = 1, 2, \ldots$. Also the matrix of the directions after the ith transformation, $D_i = \{\mathbf{d}_i(1), \ldots, \mathbf{d}_i(p)\}$, is such that $D_i = D_{i-1} P_i$.

Now it can be shown that after each transformation P_i, involving directions $\mathbf{d}_i(r)$ and $\mathbf{d}_i(s)$, $v_i(t, u) = v_{i-1}(t, u)$ where $t \neq r, s$ and $u \neq r, s$. Also it can be shown that $v_i(r, r) + v_i(s, s) = v_{i-1}(r, r) + v_{i-1}(s, s)$. Thus, as each transformation is set up so that $v_i(r, r) \geqslant v_{i-1}(r, r)$ when $v_{i-1}(r, r) \geqslant v_{i-1}(s, s)$, there must therefore be a time I where the ordering of the columns with respect to $v(r, r)$ is no longer changed by transformations P_i for $i > I$. Furthermore the size of $v_i(r, r)$ where $v_i(r, r) \geqslant v_i(s, s)$ for $s = 1, \ldots, p$ is bounded above by the largest eigenvalue of the variance–covariance matrix of the data with respect to the normalized directions $\mathbf{d}_0(1)/l_{0_1}, \ldots, \mathbf{d}_0(p)/l_{0_p}$ because the largest eigenvalue represents the maximum possible variance that can be achieved in any normalized direction that is a linear combination of the original directions (see, for example, Krzanowski and Marriott (1994)). Thus $v_i(r, r)$ where $v_i(r, r) \geqslant v_i(s, s)$ for $s = 1, \ldots, p$ must converge as i increases. By a similar argument the values of all $v_i(s, s)$, $s = 1, \ldots, p$, must also converge.

From the preceding argument it is clear that the order in which the simplicity preserving transformations are performed can be chosen in various ways.

For example one strategy is to choose at every stage the transformation for which $\max\{v_i(r, r) - v_{i-1}(r, r)\}$, $r = 1, \ldots, p$, is the greatest. Alternatively the pair of directions r, s could be chosen for which $|v_{i-1}(r, s)|$ is maximized. Using this strategy it is easy to see that this whole algorithm can be viewed as a modification of Jacobi's algorithm for evaluating eigenvectors. For further details see Krzanowski (1971).

We prefer a slightly different strategy. The choice of columns to be involved in the next transformation is restricted to those which were not affected by previous transformations. Then out of this restricted set is a pair of directions chosen so that $v_i(r, r) - v_{i-1}(r, r)$ is maximized. Only when there are no more non-trivial transformations that can be carried out are transformations on all the directions considered again. With this strategy, more than one transformation can be performed simultaneously. Additionally the direction in which $v(r, r)$ is the highest appears, in practice, often to depend equally on some or all of the original directions.

2.3. *Worked example — resistance index data*

As an example, we shall demonstrate our approach on some RI data. The RI is a measure of resistance to flow in blood-vessels. A large RI indicates that a blood-vessel is more resistant to blood flow. In a study on the ultrasound monitoring of pregnant women, RI measurements were taken on 444 women scanned between 18 and 32 weeks' gestation (Thompson *et al.*, 1999). On each woman four separate measurements of the RI from the uterine artery were recorded. These corresponded to the RI in the uterine artery on each side of the body (right and left) using two different techniques (Doppler and colour velocity imaging (CVI)). Taking $\mathbf{d}_1$, . . ., $\mathbf{d}_4$ as directions of unit length along the axes, right-side Doppler RI, left-side Doppler RI, right-side CVI RI and left-side CVI RI respectively, the corresponding variance–covariance matrix (multiplied by 100) is

$$v_0 = \begin{pmatrix} 1.552 & 0.696 & 1.258 & 0.674 \\ 0.696 & 1.528 & 0.626 & 1.298 \\ 1.258 & 0.626 & 2.543 & 0.759 \\ 0.674 & 1.298 & 0.759 & 2.663 \end{pmatrix}.$$

Now, taking $k = 0$, the first simplicity preserving transformation considers directions of the form

$$(\mathbf{d}_r + \beta_{rs}\mathbf{d}_s)/(1 + \beta_{rs}^2)$$

for $s > r$ and $\beta_{rs} = \beta_{rs}^*$ such that the variance of expression (1) is maximized. For this example, $v_1(r, r) \neq v_1(s, s)$ for $r \neq s$ and hence β_{rs}^* which maximizes the variance is not in general rational (Table 1). Consequently $b_{rs}^* \neq \beta_{rs}^*$ for any r, s.

Comparing the transformations between the different directions, overall the best simplicity preserving transformation to perform is one involving $\mathbf{d}_1$ and $\mathbf{d}_3$ with $b_{13}^* = 1$. However, a simplicity preserving transformation involving the remaining two directions $\mathbf{d}_2$ and $\mathbf{d}_4$ with $b_{24}^* = 1$ also produces an improvement in variance and thus it will also be performed simultaneously. Hence

$$P_1 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$$

and

$$v_1 = \begin{pmatrix} 3.306 & 1.377 & -0.496 & -0.055 \\ 1.377 & 3.393 & -0.007 & -0.567 \\ -0.496 & -0.007 & 0.790 & 0.078 \\ -0.055 & -0.567 & 0.078 & 0.798 \end{pmatrix}.$$

On the second stage β_{rs}^* is small for some choices of r and s, indicating that for those pairs of directions there is little scope for improvement in v_{rs}. In fact with $k = 0$ only two non-trivial simplicity preserving transformations are possible in the second sweep: one between the previously transformed $\mathbf{d}_1$ and $\mathbf{d}_2$ with $b_{12}^* = 1$ and the other between the previously transformed $\mathbf{d}_3$ and $\mathbf{d}_4$ with $b_{34}^* = 1$ (Table 2). Hence

$$P_2 = \begin{pmatrix} 1 & -2 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

Table 1. Optimal choices for β on the first sweep of the RI data

r	s	β_{rs}^*	b_{rs}^*	$v_1(r, r) \times 100$
1	2	0.983	1	2.237
1	3	1.469	1	3.306
1	4	2.121	1	2.781
2	3	2.098	1	2.662
2	4	1.529	1	3.393
3	4	1.082	1	3.362

Table 2. Optimal choices for β on the second sweep of the RI data

r	s	β^*	b^*	$v_2(r, r)\times 100$
1	2	1.032	1	4.727
1	3	−0.190	0	3.306
1	4	−0.022	0	3.306
2	3	−0.003	0	3.393
2	4	−0.209	0	3.393
3	4	1.057	1	0.871

and

$$v_2 = \begin{pmatrix} 4.727 & 0.044 & 0.563 & 0.060 \\ 0.044 & 1.972 & 0.012 & 0.501 \\ 0.563 & 0.012 & 0.871 & 0.004 \\ 0.060 & 0.501 & 0.004 & 0.716 \end{pmatrix}.$$

At the third stage there are no more non-trivial simplicity preserving transformations that can be applied and so the algorithm is over.

The overall transformation is such that

$$P = P_1 P_2 = \begin{pmatrix} 1 & -2 & -1 & 2 \\ 1 & 2 & -1 & -2 \\ 1 & -2 & 1 & -2 \\ 1 & 2 & 1 & 2 \end{pmatrix}.$$

The simple components are proportional to the columns of the matrix P. Thus, in this example, the four simple components can be written as (1, 1, 1, 1), (1, −1, 1, −1), (1, 1, −1, −1) and (1, −1, −1, 1). The variances of the normalized components can be extracted from the main diagonal of matrix v_2. So we find that the simple components explain 57%, 24%, 11% and 9% of the variance respectively. Unfortunately, although the simple components are, by construction, orthogonal, they are not in general uncorrelated. However, here the correlation between the first two simple components, which together explain 81% of the total variance, is only 0.014.

In the context of these data, the interpretation of the components is straightforward. The first simple component represents an overall measurement of the RI as it is proportional to the simple mean of all four measurements. Consequently the variance of this first component represents between-woman variability. The second component represents a contrast between the RI in the uterine arteries on the right and left sides of the body. Thus its variance (24%) represents 'location' variability within women. Similarly the third component represents a contrast between measurements of the RI using the two different techniques. Thus the variance for the third component represents the variation that is attributable to the technique used to take the measurement. Finally the fourth component represents the interaction between location and technique. It is perhaps disappointing that the variance attributed to this last component is comparable with that attributed to the third component (the technique component). However, this is in part due to a relatively high correlation (0.421) between the second and fourth components.

A comparison of the normalized simple components with the principal components (Tables 3 and 4) shows that for the RI data the approximation is quite good. The magnitude

Table 3. Simple components for the RI data

Component	*Variance*	*Cumulative variance (%)*	*Angle with corresponding exact component (deg)*	*Directions*			
$k = 0$							
1	0.0473	57.0	8.3	1	1	1	1
2	0.0197	80.8	−19.3	1	−1	1	−1
3	0.0087	91.4	−8.2	1	1	−1	−1
4	0.0072	100	19.3	1	−1	−1	1
$k = 1$							
1	0.0468	56.5	10.4	1	1	2	2
2	0.0215	82.4	−1.6	1	−1	2	−2
3	0.0092	93.5	−10.4	2	2	−1	−1
4	0.0054	100	−1.7	−2	2	1	−1
$k = 2$							
1	0.0481	58.0	1.4	3	3	4	4
2	0.0214	83.9	−3.2	8	−8	19	−19
3	0.0079	93.4	1.4	−4	−4	3	3
4	0.0055	100	3.2	19	−19	−8	8

Table 4. Principal components for the RI data

Component	*Variance*	*Cumulative variance (%)*	*Normalized directions*			
1	0.0481	58.0	0.423	0.425	0.548	0.583
2	0.0215	83.9	0.319	−0.296	0.649	−0.625
3	0.0079	93.5	0.579	0.551	−0.433	−0.415
4	0.0054	100	0.619	−0.654	−0.302	0.311

of the loadings on the variables for each of the components is close and the angle between simple components and their corresponding principal components is small.

Increasing k for the RI data shows that, for these data at least, the normalized simple components become generally closer to the exact principal component solution. This property is to be expected given the links between this approach and Jacobi's method for calculating eigenvalues and their corresponding vectors. However, this process is not monotone for all components. What is also striking is that for these data the general structure of the simple components remains despite the modest changes in k. Consequently, in this case, changing k appears merely to change the weights in the contrasts, leaving their essential interpretation untouched.

3. Another principal components analysis example — Jackson's hearing loss data

Jackson (1991) described a principal components analysis relating to hearing loss due to the normal aging process. The data consisted of hearing loss measured on 100 39-year-old males, none of whom were thought to have any particular hearing difficulties. On each subject eight measurements were taken: hearing loss in each ear at each of the four frequencies 500 Hz,

Table 5. Component loadings for the first four principal components for Jackson's hearing loss data

Frequency (Hz)	*Ear*	*Results for the following components:*			
		1	*2*	*3*	*4*
500	Left	0.952	−0.616	−0.281	−0.672
1000	Left	1.000	−0.438	0.092	−0.988
2000	Left	0.869	0.465	0.835	−0.578
4000	Left	0.667	0.922	−0.764	−0.330
500	Right	0.815	−0.751	−0.460	1.000
1000	Right	0.976	−0.451	0.052	0.762
2000	Right	0.741	0.617	1.000	0.801
4000	Right	0.603	1.000	−0.757	0.326
Variance		3.93	1.62	0.98	0.47
Cumulative variance (%)		49.1	69.3	81.5	87.4

Table 6. Component loadings for the first four simple components using $k = 0$ for Jackson's hearing loss data

Frequency (Hz)	*Ear*	*Results for the following components:*			
		1	*2*	*3*	*4*
500	Left	1	−1	0	−1
1000	Left	1	−1	0	−1
2000	Left	1	1	1	−1
4000	Left	1	1	−1	−1
500	Right	1	−1	0	1
1000	Right	1	−1	0	1
2000	Right	1	1	1	1
4000	Right	1	1	−1	1
Variance		3.86	1.59	0.98	0.46
Cumulative variance (%)		48.2	68.1	80.4	86.1
Angle with corresponding principal component (deg)		9.5	17.1	−19.2	19.7

1000 Hz, 2000 Hz and 4000 Hz. For the analysis the correlation matrix was used so that each variable would be equally weighted.

In his interpretation of the results, Jackson focused on the first four components. He identified the first component as a measure of average hearing loss noting that it loads highly and positively on all the variables (Table 5). Similarly he identified the second component as representing a contrast between high and low frequency hearing loss, the third component as primarily a contrast between hearing loss at the two higher frequencies and finally the fourth component as a contrast between the two ears.

We obtained simple components for these data after only four transformations of the correlation matrix when $k = 0$. The first four components (Table 6) clearly have the same interpretation as Jackson gave to the first four principal components. The correspondence between the simple and principal components is also obvious from the small angles between the pairs of components.

A comparison of the variances of the principal and simple components shows that little explanatory power (in terms of variance) is lost by this radical simplification. Furthermore the highest correlation between the first four simple components is only 0.151.

4. Discussion

We have introduced an algorithm that produces simple components that can be regarded as simple approximations to the principal components. This algorithm, based on Jacobi's method for evaluating eigenvectors and eigenvalues, is iterative. However, in practice we have found that it has worked quickly, coming to halt in a number of steps roughly proportional to the dimension of the data.

The motivation behind this algorithm is close in spirit to the simplification of principal components by rounding, often drastically, individual elements and the simple components sought by Hausman's branch-and-bound algorithm (Hausman, 1982). We have regarded simple components as those that can be represented by integer vectors. Furthermore we have implicitly regarded those components that are proportional to integer vectors with elements that are small as particularly simple. Consequently the algorithm is designed so that the first few components tend to be simpler than later components. However, unlike Hausman (1982), we do not explicitly restrict the magnitude of any one of the elements in any of the integer vectors.

Other definitions of simplicity have been used to motivate different strategies for the simplification and interpretation of principal components. For example, simple components could be regarded as those that are a linear combination of as few variables as possible. It is this notion of simplicity that underpins the regression strategy described in Cadima and Jolliffe (1995) and the search procedure suggested by McCabe (1984). Alternatively, simple components could be regarded as those for which the sum of the absolute values of the elements is below a given threshold (Jolliffe and Uddin, 1999). Finally one could use one of the formalized definitions of simplicity used in factor analysis such as varimax (Kaiser, 1958) or quartimax (Neuhaus and Wrigley, 1954). This leads to a rotation strategy for simplifying principal components. Usually the rotation is performed only after some of the principal components have been dropped from consideration. However, an approach is currently being developed where the rotation of components is conducted simultaneously with the extraction of components (Jolliffe and Uddin, 1998).

Within our definition of simplicity there is unfortunately no guarantee that our algorithm will produce the best simple solution. In particular, modifications of the algorithm, changing the way in which simplicity preserving transformations are put together, can result in different simple components. However, discrepancies in the simple components that are obtained do not necessarily represent a disadvantage of the approach. For us simplicity is a largely subjective concept. We have not explicitly defined a measure of simplicity, let alone quantified a trade-off between simplicity and closeness to the principal components. Instead we recognize that the utility of a particular set of simple components might genuinely vary between investigators.

Nevertheless, by focusing on the production of simple components, the investigator is at least given some feel for the extent to which a straightforward interpretation of the principal components is possible. In the RI data (Section 2.3) and Jackson's hearing loss data (Section 3), in common with other examples that we have tried, the simple components produced by the algorithm with $k = 0$ provided good approximations, indicating that easily interpretable structure was there to be found.

In contrast, in a data set where simple structure has proven difficult to detect in the past (the 'pit-prop' data (Jeffers, 1967)) the simple components algorithm struggles. The simple components produced when $k = 0$ are not as close to their corresponding principal components (Table 7) as was seen in Sections 2.3 and 3. Nor do higher values of k produce components

Table 7. Component loadings for the first five simple components using $k = 0$ for Jeffers's pit-prop data

Variable	*Results for the following components:*				
	1	*2*	*3*	*4*	*5*
Top diameter	1	2	2	−133	620571
Length	1	2	2	−133	620571
Moisture content	0	1	−9	603	−745121
Specific gravity	0	1	−9	601	−744591
Oven dry specific gravity	1	−2	−5	79	3491021
Annual rings (top)	1	0	−11	−333	−2253059
Annual rings (bottom)	1	0	−9	−273	−2268959
Maximum bow	1	0	8	250	−1236902
Bow distance	1	1	1	−68	603346
Knot whorls	1	0	8	224	−240664
Clear prop length	0	1	0	20	11701220
Average knots or whorls	−1	1	−7	−308	666946
Average knot diameter	−1	2	3	−79	−3491021
Variance	3.67	2.40	1.81	1.05	1.02
Cumulative variance (%)	28.2	46.7	60.6	68.7	76.5
Angle with corresponding principal component (deg)	27.4	43.4	−34.3	73.9	−71.9

with a better balance between simplicity and high variance. However, some interpretation is still immediate from the simple components with $k = 0$. The first simple component loads equally on top diameter, length, number of annual rings at the top and bottom, the bow and the number of whorls like the first principal component and it also picks up the contrast between them and the average knots or whorls and diameter of knots. More strikingly the loadings for the top diameter and the length of the prop are identical for the first five simple components, indicating that those two measurements essentially measure the same thing.

Our algorithm can be applied to any symmetric matrix where there is an interest in maximizing the elements down the main diagonal and hence it could be applied to techniques such as principal co-ordinates analysis. However, its extension to other multivariate techniques based on eigenvectors is not immediate. For example, in discrimination, orthogonality is defined with respect to the within-groups variance–covariance matrix W. This makes it difficult to define an initial set of simple vectors that are orthogonal with respect to W. Consequently there may be no initial simplicity for the simplicity preserving transformation described in Section 2.1 to preserve. However, work is in progress to overcome this difficulty by relaxing the orthogonality restriction.

Similarly problematical are the related techniques of canonical correlation and correspondence analysis. Here again orthogonality is defined with respect to matrices other than the identity matrix, again raising the problem of defining an initial set of simple orthogonal vectors. Furthermore, in these techniques, variables are split into two distinct groups with the optimality criterion (maximal correlation) depending on two linked linear combinations of variables. Thus the algorithm as described here would also have to be adapted to be in line with that optimality criterion. Finally, in principle, the algorithm could be applied without change to factor analysis by embedding it within the principal factoring approach. However, there is no guarantee that the resulting combined algorithm will converge. Indeed, already our experience with one example is that the combined algorithm oscillated between two different solutions instead of converging to a single solution. This suggests that a different strategy is required before the principles underlying our algorithm can be applied to factor analysis.

Acknowledgements

I am grateful to Chris Jones for many fruitful discussions and for interesting me in this topic. I am also grateful to Dr M. O. Thompson, Milton Keynes General Hospital, for allowing me to use the RI data.

References

Cadima, J. and Jolliffe, I. T. (1995) Loadings and correlations in the interpretation of principal components. *J. Appl. Statist.*, **22**, 203–214.

Hausman, R. E. (1982) Constrained multivariate analysis. In *Optimization in Statistics* (eds S. H. Zanakis and J. S. Rustagi), pp. 137–151. Amsterdam: North-Holland.

Jackson, J. E. (1991) *A Users Guide to Principal Components Analysis*. New York: Wiley.

Jeffers, J. N. R. (1967) Two case studies in the application of principal component analysis. *Appl. Statist.*, **16**, 225–236.

Jolliffe, I. T. and Uddin, M. (1998) The simplified component technique: an alternative to rotated principal components. Unpublished. University of Aberdeen, Aberdeen.

———(1999) A modified principal component technique based on the lasso. Unpublished. University of Aberdeen, Aberdeen.

Kaiser, H. F. (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**, 187–200.

Krzanowski, W. J. (1971) The algebraic basis of classical multivariate methods. *Statistician*, **20**, 51–61.

Krzanowski, W. J. and Marriott, F. H. C. (1994) *Multivariate Analysis, Part 1*. London: Arnold.

McCabe, G. P. (1984) Principal variables. *Technometrics*, **26**, 137–144.

Neuhaus, J. and Wrigley, C. (1954) The quartimax method: an analytical approach to orthogonal simple structure. *Br. J. Statist. Psychol.*, **7**, 81–91.

Thompson, M. O., Vines, S. K. and Harrington, K. (1999) Assessment of blood volume flow in the uterine artery; the influence of arterial distensibility and waveform abnormality. *Ultsound Obstet. Gyn.*, **14**, suppl. 1, 71.